Spring 2017

Validity of energy expenditure for the Garmin Vivofit during a week of physical activity

Kristen Webb
James Madison University

Validity of Energy Expenditure for the Garmin Vivofit during a Week of Physical Activity

---

An Honors College Project Presented to

the Faculty of the Undergraduate

College of Kinesiology

James Madison University

---

by Kristen Renee Webb

April 2017

Accepted by the faculty of the Department of Kinesiology, James Madison University, in partial fulfillment of the requirements for the Honors College.

FACULTY COMMITTEE:

Project Advisor: Trent A. Hargens, Ph.D
Exercise Science Professor, Kinesiology

Reader: Christopher J. Womack, Ph.D
Department Head, Kinesiology

Reader: Elizabeth S. Edwards, Ph.D
Exercise Science Professor, Kinesiology

HONORS COLLEGE APPROVAL:

Bradley R. Newcomer, Ph.D.,
Dean, Honors College

PUBLIC PRESENTATION

This work is accepted for presentation, in part or in full, at the Kinesiology Symposium on 4/20/17.

Table of Contents

Acknowledgements

First and foremost, I would like to thank my project advisor, Dr. Hargens for all of his guidance and help throughout the process of completing my Honors Capstone Project. None of this would have been possible without his knowledge and expertise to help me along the way and to encourage me to create the best product possible. This process has been an incredible experience that I believe will help me in my future endeavors due to the large amount of knowledge received from Dr. Hargens.

I would also like to thank Dr. Edwards and Dr. Womack for helping with my Honors Capstone Project by taking time out of their busy schedules to be readers for my project. Their advice and guidance throughout the project was critical in making my project a success.

Finally, I would like to thank the JMU Kinesiology Department for allowing me to complete this research under Dr. Hargen's supervision and in its facilities. I feel glad to have had the opportunity to conduct research within the department and to have had this experience during my undergraduate career.

List of Tables and Figures

Abstract

Purpose To determine the accuracy of energy expenditure measured by the Garmin Vivofit during a 7 day period of daily living in comparison to a research grade accelerometer, Actigraph.

Methods Participants wore a research grade accelerometer (Actigraph GT3X+) and a Garmin Vivofit device for 7 straight days, 24 hours a day. Paired *t*-tests were used to examine mean differences in 7-day mean energy expenditure and step values obtained between the devices, as well as the mean energy expenditure values for each day. Pearson correlation was used to assess the relationship of energy expenditure between the devices. Statistical significance was set *a priori* at $p < 0.05$.

Results There was no significant difference between the Actigraph and the Vivofit for mean total daily energy expenditure. The Vivofit reported a significantly lower mean daily active energy expenditure (271.13 vs 513.02 kcals, for Vivofit vs. Actigraph, respectively, $p < 0.001$). The Vivofit reported significantly higher mean daily step averages (13,412.00 vs. 11,307.58, for Vivofit vs. Actigraph, respectively, $p = 0.01$). Paired t-test analysis of mean active kilocalories concluded a significance difference on all 7 days between the devices. There was no significant difference between the devices for mean total kilocalories for any day. Pearson correlations were highly significant between the devices for both total and active kilocalories ($r = 0.99$, $p < 0.001$).

Conclusion The Vivofit total energy expenditure results matched expected results of having no significant difference between the devices. Active energy expenditure and step results were unexpected with the Vivofit underestimating daily active energy expenditure and overestimating daily step averages as compared to the Actigraph.

Keywords: Actigraph, Garmin Vivofit, energy expenditure

Chapter I

Introduction

Commercial physical activity monitors, like the Garmin Vivofit, have become very popular devices in recent years. A commercial physical activity monitor is a device based off of accelerometer technology which tracks different variables of fitness. Prior to the creation of these commercial devices, accelerometers were used primarily in a research environment due to the high expense and complex nature of the device. However, in recent years, accelerometers have become accessible and easy to access allowing companies to commercialize these products [5]. These commercial devices have influenced consumers through their ability to track level of fitness, physical health, and weight maintenance [7]. Individuals can see a variety of measurements like step count, caloric expenditure, and even heart rate, just to name a few, allowing for quick access to fitness information that was not traditionally available to the general population. The ability for people to receive this feedback, from a simple wrist or hip monitor, has increased the use of these devices. Commercial wrist monitors include devices such as the FitBit Flex, Nike FuelBand, Jawbone Up, Misfit Shine, Withings Pulse, and Garmin Vivofit. Commercial hip monitors include devices such as the Misfit Shine, FitBit Zip, FitBit One, and FitBit Ultra. The desire for these devices is so high that commercial activity monitor sales was expected to reach 45 million devices by this year [4]. With a surge in popularity of these devices in the population, assessment of their respective validity is important.

The commercial physical activity monitors can be used as a measure of daily physical activity as well as structured exercise. Currently, there has been some research conducted on the validity of certain devices with respect to step count. One study, in particular, focused on determining the accuracy of step count measurements of multiple devices such as the FitBit Flex, Jawbone UP, Nike FuelBand, and Garmin Vivofit. The researchers of this study performed a 200, 500, and 1,000 step test where steps were counted by tally along with the devices to determine the level of accuracy. This study revealed, in particular, that the Garmin Vivofit was 97.01% accurate in calculating step count. The FitBit Flex was 80.43%, Jawbone UP was 82.51%, and Nike FuelBand was 95.64% accurate, respectively; showing that device accuracy is dependent on the fitness device [4].

Another study focused on analyzing the accuracy of ten different commercial devices by conducting a treadmill test as well as a free living test. For the treadmill test, participants walked at a speed of 4.8 km/h for 30 minutes while wearing all of the devices. The Optogait system, a light sensor on the treadmill that tracks step count with the movement of the participant, was used for comparison during the treadmill test. For the free living test, the ActivPAL device was used as the standard for comparison. For the treadmill test, the mean percentage error for the FitBit Flex was 5.7%, Jawbone UP was 1%, Nike FuelBand was 18%, Misfit Shine was 0.2%, Withings Pulse was 0.5%, and FitBit Zip was 0.3%. For the free living test, the mean percentage error for the FitBit Flex was 3.7%, Jawbone UP was 1.4%, Nike FuelBand was 24%, Misfit Shine was 1.1%, Withings Pulse was 7.9%, and FitBit Zip was 1.2%. To rate validity of the devices during the treadmill test, a device with a percentage error higher than 1% was considered to have low accuracy for this experiment. Therefore, the FitBit Flex and Nike FuelBand were unreliable sources of step count during the treadmill test. On the other hand, a percentage error larger than 10% was considered as a standard for low step count accuracy during the free living test. By this standard, the Nike Fuel Band was considered an unreliable source of step count for the free living test. This data shows reliability of step count, in particular, is dependent on type of physical activity with the structured exercise test having a higher overall accuracy in comparison to daily living activities [6].

Storm et al. evaluated the accuracy of step count for seven different physical activity monitors by testing a larger variety of activities at three different speeds. The testing protocol for this study included indoor walking, outdoor walking, walking down steps, and walking up steps for a total of 11 minutes at a slow pace, self-selected pace, and fast pace where an OPAL sensor was used as the means for comparison. The study concluded that the FitBit One and Nike FuelBand underestimated step count with a mean percent error of less than 2.6% for all speeds for the FitBit One and less than approximately 35% for all speeds for the Nike FuelBand, where percent error decreased with an increase in speed. The Jawbone UP accurately estimated step count and had a mean percent error of less than 10% for all speeds, where percent error decreased with an increase in speed. Overall, this study found that slower speeds during any activity causes an increase in percent error for the particular physical activity monitors. Therefore, the speed of activities influences the level of accuracy of step count measurements [13].

Nelson et al. validated step count accuracy for the FitBit One, FitBit Zip, FitBit Flex, and the Jawbone UP during three types of activities considered sedentary, household, and ambulatory. An Omron device was used as well as actually counting the steps for a standard of comparison to the four devices studied. To determine the level of accuracy, mean percent errors were calculated for each device for both the household and ambulatory activities. For the household activities, the percent error for the FitBit One was 71%, FitBit Zip was 70%, FitBit Flex was 58%, and Jawbone UP was 54%, thus showing low accuracy for all devices. Furthermore, for the ambulatory activities, the percent error for the FitBit One was 2%, FitBit

Zip was 3%, FitBit Flex was 6%, and Jawbone UP was 6% showing high accuracy in comparison to the household activities. This study showed that step count accuracy, similar to other findings, depends on activity and that household activities seem to elicit the least accurate and most variable step count results [9].

Takacs et al. evaluated the FitBit One for step count accuracy during a period of treadmill walking at five different speeds as well as different locations for placing the device. At the speeds of 0.9 m/s, 1.12 m/s, 1.33 m/s, 1.54 m/s, and 1.78, the intra-class correlation was 0.99, 1.00, 0.99, 0.99, and 0.95, respectively, illustrating that the FitBit One had a very high reliability for step count during the treadmill walking protocol. Also, for all of the speeds, all of the step counts were no more than 4 steps apart, therefore, further illustrating the device's high level of accuracy [14].

Furthermore, Tully et al. analyzed the step count accuracy of the FitBit Zip device during a week of daily living where the device was compared to a research grade accelerometer, Actigraph, as well as a pedometer. The researchers compared the median number of steps per day of each device to determine the level of accuracy. The FitBit Zip recorded 7,477 steps per day, the research accelerometer recorded 6,774 steps per day, and the pedometer recorded 7,532 steps per day showing, suggesting a low validity for the FitBit Zip [15]. From the multiple studies described above, research has evaluated the validity of these devices on step count in a variety of situations. However, more devices and activities need to be studied to completely cover the range of commercial monitors on the market.

Beyond analysis of step count, energy expenditure during structured exercise and daily living has been studied to a small extent, but no research thus far has involved the Garmin Vivofit. One study involving treadmill exercise found that the FitBit Tracker and FitBit Ultra measured a lower energy expenditure than what actually occurred during the exercise performances. Subjects completed three treadmill exercises at 3.5 mph with no incline, 3.5 mph and 5% incline, and 5.5 mph with no incline in order to simulate walking, walking uphill, and running. Subjects also completed stair stepping as part of the testing protocol. Correlations between each device and indirect calorimetry were used to determine the validity of the devices for each mode of exercise performed. During walking, the FitBit Tracker had a correlation of 0.70 and the FitBit Ultra had a correlation of 0.83 to indirect calorimetry. For the walking uphill bout of exercise, the correlation values were 0.72 for the FitBit Tracker and 0.81 for the FitBit Ultra when compared to indirect calorimetry. For the running exercise bout, the correlation values were 0.56 for the FitBit Tracker and 0.87 for the FitBit Ultra compared to indirect calorimetry. Finally, the stepping bout of exercise had correlation values of 0.18 for the FitBit Tracker and 0.58 for the FitBit Ultra compared to indirect calorimetry. From the correlation values, validity for energy expenditure was fairly low, with the FitBit Ultra showing a higher overall accuracy [10].

Dannecker et al. evaluated the accuracy of a FitBit Tracker, worn at the waist, during a test of either treadmill walking, cycling, or stepping. A lower estimated energy expenditure (362.8 ± 18.9 kcal) was observed compared to indirect calorimetry (499 ± 23.8 kcal). Similar findings were also observed by Noah et al. [2].

In contrast, Diaz et al. concluded that the FitBit One and the FitBit Flex were fairly accurate in estimating energy expenditure. A staged treadmill protocol was implemented at 1.9 mph, 3 mph, 4 mph, and 5.2 mph in order to simulate walking at different speeds as well as running. The exception to the study's findings was the higher estimated energy expenditure values for the wrist technology, FitBit Flex, during the 3 to 4 mph walking exercise bouts.

Researchers used within-participant correlations to determine the level of accuracy of energy expenditure for each device in comparison to measured energy expenditure through indirect calorimetry. The FitBit One and the FitBit Flex had high correlation values of 0.86 and 0.88, respectively, across all stages showing a fairly high level of accuracy. Thus, there is contradicting results between the varieties of FitBit devices as seen with the FitBit One, a hip worn device, and the FitBit Flex, a wrist worn device. This shows that there can also be variability of accuracy for energy expenditure in a single physical activity monitor brand as well as between different devices [3].

Furthermore, one study has been conducted comparing estimated energy expenditure from commercial physical activity monitors (Jawbone UP, Nike FuelBand, and Fitbit Ultra), during a variety of forms of exercise, including walking and running on a treadmill, exercise on an elliptical, and agility drills. For the treadmill walking protocol, the Jawbone UP measured 123 kcal, the Nike FuelBand measured 107 kcal, the FitBit Ultra measured 111 kcal, and the Adidas MiCoach measured 146 kcal in comparison to the 109 kcal measured from indirect calorimetry. For the treadmill running protocol, energy expenditure values were 288 kcal for the Jawbone UP, 275 kcal for the Nike FuelBand, 230 kcal for the FitBit Ultra, and 261 kcal for the Adidas MiCoach in comparison to 240 kcal measured through indirect calorimetry. For the elliptical exercise bout, energy expenditure values were 161 kcal for the Jawbone UP, 118 kcal for the Nike FuelBand, and 154 kcal for the FitBit Ultra in comparison to the 161 kcal measured by indirect calorimetry. Finally, for the agility drills, energy expenditure values were 63 kcal for the Jawbone UP, 77 kcal for the Nike FuelBand, 75 kcal for the FitBit Ultra, and 36 kcal for the Adidas MiCoach in comparison to 90 kcal measured by indirect calorimetry. Overall, research found that performance of activities that strayed from "normal ambulation" resulted in more inaccurate values for energy expenditure [12].

Beyond structured exercise test protocols, Lee et al evaluated energy expenditure for a larger variety of activities which are common to daily living such as, sedentary time, walking, running, cycling, and stair climbing. Out of all of the commercial devices studied, the lowest percent error was approximately 9.3% for the BodyMedia Fit device, an armband research accelerometer which measures energy expenditure through heat, and the highest percent error was 23.5% for the Basis band, a wristband accelerometer which also measures energy expenditure through heat. The other devices within the study had percentage errors of 10.1% for the FitBit Zip, 10.4% for the FitBit One, 12.2% for the Jawbone UP, and 13% for the Nike FuelBand [7]. This study suggested that the accuracy of energy expenditure is affected by the type of daily activities.

Ferguson et al. concluded that energy expenditure values by seven different commercial devices, in general, were lower than the accepted reference device mean value of 3,005 kcal during a 24 hour period of daily living. The BodyMedia device, a research grade accelerometer, was used as the reference device for comparison to the commercial devices studied for energy expenditure. The validity coefficients for the devices were 0.74 for the Jawbone UP, 0.76 for the FitBit One, 0.79 for the Misfit Shine and Withings Pulse, and 0.81 for the FitBit Zip. All of the devices were moderately accurate in analyzing energy expenditure relative to the reference device. The study also concluded that energy expenditure measurements were the least accurate between the devices studied in comparison to other variables, such as step count [5].

Finally, one study focused specifically on different daily living activities. This study evaluated the FitBit One, FitBit Zip, FitBit Flex, and the Jawbone UP. Indirect calorimetry was used as the criterion measurement. For the sedentary activities, the measured energy expenditure values were 28.7 kcal for the FitBit One, 29.9 kcal for the FitBit Zip, 28.7 kcal for the FitBit Flex, and 28.3 kcal for the Jawbone UP in comparison to 30.8 kcal measured by indirect calorimetry. For the household activities, the energy expenditure values were 52.6 kcal for the FitBit One, 47.2 kcal for the FitBit Zip, 66.5 kcal for the FitBit Flex, and 47.8 kcal for the Jawbone UP in comparison to the indirect calorimetry value of 71.7 kcal. Finally, for the ambulatory activities, the energy expenditure values were 143.2 kcal for the FitBit One, 172.3 kcal for the FitBit Zip, 163.8 kcal for the FitBit Flex, and 145.2 kcal for the Jawbone UP in comparison to 123.5 kcal from indirect calorimetry [9]. From the results of this study, energy expenditure measurements are clearly variable between devices and types of activities with sedentary activities being the most accurate in measuring energy expenditure in comparison to household and ambulatory activities.

From the various studies listed involving energy expenditure, much of the research is specific to FitBit technology, Jawbone UP, and Nike FuelBand devices. In contrast, other devices such as the Garmin Vivofit have not been evaluated. Overall, the level of accuracy between all of the devices studied is extremely variable with no clear conclusion on which devices provide the best overall energy expenditure measurement. However, the FitBit devices, in general, appear to be the most accurate for all of the different activities studied whether in a lab environment or free living environment, including running and walking. The Nike FuelBand device appears to be the most accurate at lower levels of intensity such as seen with a higher accuracy during walking and a lower accuracy during running. Furthermore, the Jawbone UP device, in general, appears to be collectively the least accurate when compared to the Nike FuelBand and FitBit devices over the variety of activities studied.

Possible limitations with determining a clear list of the most accurate devices include the inability for most of these devices to track incline changes as well as varying movements and intensities during a given activity efficiently and effectively. Furthermore, the majority of energy expenditure research has focused on specific tasks, mostly running and/or walking. The type of exercises studied is very limiting in understanding the level of energy expenditure during daily living physical activity because on an average day individuals will perform more activities than just walking and running that expend calories. Also, the research listed above has some contradicting results which requires further exploration in order to have clear results explaining which devices are the most accurate in calculating energy expenditure. Furthermore, much of the research listed above focuses on controlled exercise testing as opposed to free living testing. Therefore, there is a need for research in examining more devices as well as studying the free living setting.

Purpose

The purpose of this study was to determine the accuracy of energy expenditure measured by the Garmin Vivofit device during a seven day period of daily living in comparison to the energy expenditure results measured by a research grade accelerometer (Actigraph). The results of energy expenditure from the Garmin Vivofit device and the research grade accelerometer, were expected to be similar.

Table 1. Overview of Percent Accuracies for Steps from Past Studies

Device	Basic Step Test	Treadmill Walking-Steps	Free Living-Steps	Household Activities-Steps	Ambulatory Activities-Steps	**Averages**
Adidas MiCoach	-	-	-	-	-	-
FitBit Flex	80.43%	94.3%	96.3%	42%	94%	81.4%
FitBit One	-	-	97.4%	29%	98%	74.8%
FitBit Tracker	-	-	-	-	-	-
FitBit Ultra	-	-	-	-	-	-
FitBit Zip	-	99.7%	89.6-99.9%	30%	97%	79.1-81.7%
Garmin Vivofit	97.01%	-	-	-	-	97.01%
Jawbone UP	82.51%	99%	90-98.6%	46%	94%	82.3-84%
Misfit Shine	-	99.8%	98.9%	-	-	99.4%
Nike FuelBand	95.64%	82%	65-76%	-	-	80.9-84.5%
Withings Pulse	-	99.5%	92.1%	-	-	95.8%

*Takacs et al. is not included in this data due to insufficient percent accuracy values.

Table 2. Overview of Percent Accuracies for Energy Expenditure from Past Studies

Device	Treadmill Walking-EE	Treadmill Walking with Incline-EE	Treadmill Running-EE	Stair Stepping-EE	Elliptical-EE	Agility Drills-EE	Free Living-EE	Sedentary Activities-EE	Household Activities-EE	Ambulatory Activities-EE	**Averages**
Adidas MiCoach	66.1%	-	91.3%	-	-	40%	-	-	-	-	65.8%
FitBit Flex	-	-	-	-	-	-	-	93.2%	92.7%	67.4%	84.4%
FitBit One	-	-	-	-	-	-	89.6%	93.2%	73.4%	84%	85.1%
FitBit Tracker	72.7-90%	58.6%	88.5%	55.7-72.7%	-	-	-	-	-	-	68.9-77.5%
FitBit Ultra	89-98.2%	58.9%	90.2-95.8%	55.7%	95.7%	83.3%	-	-	-	-	78.8-81.3%
FitBit Zip	-	-	-	-	-	-	89.9%	97.1%	65.8%	60.5%	78.3%
Garmin Vivofit	-	-	-	-	-	-	-	-	-	-	-
Jawbone UP	87.2%	-	80%	-	100%	70%	87.8%	91.9%	66.7%	82.4%	83.3%
Misfit Shine	-	-	-	-	-	-	-	-	-	-	-
Nike FuelBand	98.2%	-	85.4%	-	73.3%	85.6%	87%	-	-	-	85.9%
Withings Pulse	-	-	-	-	-	-	-	-	-	-	-

*Diaz et al. and Ferguson et al. is not included in this data due to insufficient percent accuracy values.

Chapter II

Methods

For this study, we recruited 10 participants on a voluntary basis. The subjects were 18 years of age or older as well as free of any limitations that may inhibit the individual's health and/or performance of an exercise test in order to participate. Subject characteristics were normally distributed, with the exception of there being a 4:1 ratio of females to males (Table 1). Individuals with cardiovascular, pulmonary, or metabolic disease were not allowed to participate. Individuals were recruited from the James Madison University campus as well as the surrounding Harrisonburg community. The subjects were informed of the procedures of the study and provided their informed consent to participate. The procedures of the study were authorized by the James Madison University Institutional Review Board before any testing occurred.

Subjects attended and partook in two separate visitations, 8 to 10 days apart, in the Human Performance Lab. At the first visit, body weight and height were measured with a physician's scale and wall mounted stadiometer, respectively. Weight was measured to the nearest 0.1 kg, and height was measured to the nearest 0.1 cm. Percent body fat and percent lean mass were obtained via Dual Energy X-ray Absorptiometry to ensure the most accurate measurements of energy expenditure from the devices.

Each participant was given a research grade accelerometer, known as the Actigraph GT3X+, as well as a Garmin Vivofit device. The Actigraph was worn on the hip in the midline of the right leg and the Garmin Vivofit was worn on the non-dominant wrist until the individual's next visit. At night, participants were instructed to move the Actigraph to their wrist. Participants were told to wear both devices for seven straight days, twenty-four hours a day except for times when the devices could be exposed to water. Following the seven days of wearing the devices, subjects returned the devices. As part of a larger study, data collection and analysis focused on the energy expenditure values recovered from the two devices following the seven days of wear by each participant.

This study had one within subject factor (research grade accelerometer energy expenditure and commercial device energy expenditure). Intra-class correlations were utilized for the analysis of the energy expenditure between the research grade accelerometer and the Garmin Vivofit. Paired *t*-tests were also used to determine significant differences in the energy expenditures between the Actigraph and the Garmin Vivofit. Finally, Bland-Altman plots were used in order to determine the level of agreement between the Actigraph and the Garmin Vivofit in measuring energy expenditure.

Chapter III

Manuscript

Introduction

Commercial physical activity monitors, like the Garmin Vivofit, have become very popular devices in recent years. A commercial physical activity monitor is a device that is based off of accelerometer technology which tracks different variables of fitness such as, step count, calories burned, and heart rate. Prior to the creation of commercial devices, accelerometers were used primarily in research environments due to the high expense and complex nature of the device. However, in recent years, accelerometers have become more accessible, allowing companies to commercialize these products [3]. Commercial devices have influenced consumers to track level of fitness, physical health, and weight maintenance [4]. The ability to receive physical activity feedback, from a simple wrist or hip monitor, has increased demand for these devices. With a surge in popularity of commercial devices, assessing the validity of these devices is of high practical importance.

Overall, the level of accuracy between all of the devices studied thus far is extremely variable with no clear conclusion on which devices provide the best overall energy expenditure measurement. However, the FitBit devices, in general, appear to be the most accurate for all of the different activities studied whether in a lab environment or free living environment, including running and walking. The Nike FuelBand device appears to be the most accurate at lower levels of intensity such as seen with a higher accuracy during walking and a lower accuracy during running. Furthermore, the Jawbone UP device, in general, appears to be collectively the least accurate when compared to the Nike FuelBand and FitBit devices over the variety of activities studied.

Possible limitations with determining a clear list of the most accurate devices is due to the inability for most of these devices to track varying movements and intensities during a given activity efficiently and effectively [4]. Furthermore, the environment studied is very limiting in understanding the level of energy expenditure during daily living physical activity because on an average day individuals could be performing activities that require the use of different muscles that may not be tracked appropriately by the devices [4]. This limitation is seen in the research in which the least accurate results for the majority of devices is during the daily living tests.

Much of the current energy expenditure research is specific to FitBit technology, Jawbone UP, and Nike FuelBand devices which leaves room for analysis of other devices such as the Garmin Vivofit which has no current research in the area of energy expenditure. Furthermore, research conducted in a free living environment is limited with more studies involving testing in a controlled, lab environment. Therefore, the purpose of this study was to determine the accuracy of energy expenditure measured by the Garmin Vivofit device during a seven day period of daily living in comparison to the energy expenditure results measured by a research grade accelerometer, known as the Actigraph, in order to address these two main areas of limited research.

Methods

Subjects

10 participants, 18 years of age or older, on a voluntary basis from the James Madison University campus participated in this study. Subject characteristics were normally distributed, with the exception of there being a 4:1 ratio of females to males (Table 1). Subjects were free of any limitations that may inhibit the individual's health and/or performance of an exercise test.

Individuals with cardiovascular, pulmonary, or metabolic disease were not allowed to participate. Subjects were informed of the procedures of the study and provided their informed consent to participate. The procedures of the study were authorized by the James Madison University Institutional Review Board before any testing occurred.

Procedures

Subjects partook in two separate visits, 8 to 10 days apart, in the Human Performance Lab. For the first visit, body weight and height were measured with a physician's scale and wall mounted stadiometer, respectively. Percent body fat and percent lean mass were obtained via Dual Energy X-ray Absorptiometry. Following these measurements, each participant was given a research grade accelerometer (Actigraph GT3X+, ActiGraph Corp., Pensacola, FL), as well as a Garmin Vivofit device. As this was a part of a larger study, participants were told to wear both devices for 7 straight days, 24 hours a day (including sleep) except for times when the devices could be exposed to water. Following the 7 day wear period, devices were returned to the Human Performance Laboratory for download and analysis.

Statistical Analysis

All statistical analyses were conducted using SPSS statistical package (version 24.0). Mifflin-St. Jeor method was used to calculate resting metabolic rate for the Actigraph [5]. Paired *t*-tests were used in order to examine mean differences in the 7 day mean energy expenditure and step values obtained between the Actigraph and the Garmin Vivofit, as well as the mean energy expenditure values for each day. Pearson correlation was used to confirm the relationship of energy expenditure between the Actigraph and the Garmin Vivofit. Statistical significance was set *a priori* at $p < 0.05$.

Results

There was no significant difference between the Actigraph and the Vivofit for mean total daily energy expenditure for a seven day period of free living (Figure 1). However, The Vivofit reported a significantly lower mean daily active energy expenditure compared to the Actigraph (271.13 vs. 513.02 kcals, for Vivofit vs. Actigraph, respectively, $p < 0.001$) (Figure 2). In contrast, the Vivofit reported significantly higher mean daily step averages when compared to the Actigraph (13,412.00 vs.11,307.58 kcals, for Vivofit vs. Actigraph, respectively, $p = 0.01$) (Figure 3). Furthermore, mean active kilocalories was lower ($p < 0.05$) for the Vivofit for each day during the seven day period of free living (Figure 5). There was no significant difference between the Actigraph and the Vivofit for mean total kilocalories for any day during the seven day period of free living, with Sunday trending toward significant ($p = 0.054$) (Figure 4). Pearson correlations were highly significant between the Actigraph and the Vivofit for both total (Figure 6) and active kilocalories ($r = 0.99$, $p < 0.001$).

Discussion

As one of the first studies to analyze energy expenditure of the Garmin Vivofit for a seven day period of free living, results from this study concluded there was no significant difference between the Actigraph and the Vivofit for mean total daily energy expenditure, whether looking at averages for the week as a whole or for each day. However, there was a significant difference between the two devices in measuring mean active energy expenditure. In addition, the current study showed a significant difference in step data measured by both devices.

Although there was no significant difference between the Actigraph and the Vivofit in measuring mean total daily energy expenditure, past studies of free living have found similar results in that the energy expenditure values were similar to the control device. Ferguson et al. studied several devices, including the Misfit Shine, Jawbone UP, Withings Pulse, Fitbit Zip, and Fitbit One, which all showed a moderate to strong correlation in measuring total energy expenditure values when compared to two reference devices, BodyMedia SenseWear and Actigraph [3]. Additionally, there was no significant difference between the commercial devices and control devices when measuring total energy expenditure for the 24 hours of free living tested [3]. Furthermore, Lee et al., analyzed total energy expenditure for a number of devices, including a BodyMedia FIT device worn on the arm, Fitbit One and Fitbit Zip devices worn at the waist, and Jawbone Up and Basis B1 Band devices worn on the wrist. Total energy expenditure measurements for these devices were compared to the Actigraph in a testing environment which focused on performing different types of activities similar to ones performed in a daily living environment. Although this study was not performed under free living conditions, the activities performed in the study worked to imitate daily living activities. They found that total energy expenditure from the commercial devices matched fairly close to the Actigraph value of 326.2 kcals. The commercial monitor averages ranged betweem 271.1 kcals to 370.1 kcals, which is similar to Ferguson et al. as well as the present study [4].

Potential reasons for the matching results of mean total energy expenditure during activity in a free living setting could be due to the similarity of device technology between the two devices. Specifically, the Garmin Vivofit has triaxial accelerometer technology which would calculate total energy expenditure in a similar manner as the Actigraph. Also, both devices measure similar values, such as energy expenditure, steps, and sleep, potentially showing close relation in how the devices measure total energy expenditure by including both physical activity and sleep data toward those measurements. Furthermore, both devices can only estimate resting energy expenditure and most likely estimate total energy expenditure in similar manners. For the Actigraph device, we calculated resting metabolic rate with the use of the Mifflin-St. Jeor method and added this result to the active energy expenditure calculated by the Actigraph itself to get the total energy expenditure values. Since the results between the Actigraph and the Vivofit are similar, the method of estimation of RMR by the Garmin Vivofit is most likely very similar to the Mifflin-St. Jeor method used for the Actigraph [5]. These findings confirm our expectations in that calculation of total energy expenditure is very similar between a research grade accelerometer, the Actigraph, and the Garmin Vivofit, specifically during a seven day period of free living. These results further prove reliability of the Garmin Vivofit in calculating total energy expenditure.

While total energy expenditure appeared to agree between the two devices, there was a significant difference between the Actigraph and the Garmin Vivofit in measuring mean daily active energy expenditure, with the Garmin Vivofit underestimating active energy expenditure. Dannecker et al. suggests the wear location of the device causes variability in energy expenditure measurements due to the varying intensities of activities not matching the actual expenditure by the individual. They suggest that hip worn commercial devices are potentially more inaccurate in this area than wrist worn devices [1]. Furthermore, Lee et. al suggests that the free living environment also contributes to the significant difference in active energy expenditure due to including a higher proportion of activities utilizing upper body muscles, in comparison to lab tests. In summary, they suggest that commercial monitors are more biomechanically inaccurate in a free living environment due to the variability in body movements [4]. Overall, the significant underestimation of the Vivofit emphasizes how its technology should be improved to meet the level of the Actigraph, a research grade accelerometer. To date, no study has examined active energy expenditure measurements during a period of free living. However, a study by Nelson et al. exemplifies how energy expenditure calculations can vary between types of activities. Energy expenditure measured by the, Fitbit One, Fitbit Zip, and Jawbone UP24, all significantly underestimated household activity energy expenditure in comparison to indirect calorimetry. Furthermore, the three devices listed above as well as the Fitbit Flex also significantly overestimated ambulatory activity energy expenditure. The significant difference in energy expenditure measuring between the devices and indirect calorimetry during a variety of activities could exemplify a potential reason for the significant difference, as suggested by Nelson et al, due to the fact that a wrist worn device may be less accurate than a hip worn device in measuring active energy expenditure [6]. For these reasons, devices, like the Garmin Vivofit, are potentially inaccurate when measuring active energy expenditure in that the intensity of physical activity fluctuates daily based on what the person is doing every second of every day. This, for example, is an important factor for anyone trying to track weight loss by counting caloric intake and active energy expenditure output. If active energy expenditure accuracy fluctuates by the intensity and type of physical activity, individuals will be getting false results of how many calories they are truly burning compared to their food intake, thus negatively affecting weight loss progression. In this case, the Garmin Vivofit's technology may be less advanced than the Actigraph when it comes to specifically measuring energy expenditure during periods of different types and levels of physical activity. Further research of how intensity and environment affects the Garmin Vivofit's active energy expenditure calculations would need to be conducted to confirm this notion.

Finally, there was a significant difference in step count measurements between the Actigraph and the Vivofit. Paradoxically, the Garmin Vivofit significantly overestimated step counts during the seven day period of free living, while underestimating active energy expenditure. These step results were consistent with a study conducted by Tully et al. where they determined that the FitBit Zip technology also overestimated step counts in comparison to the Actigraph during a seven day period of free living [7]. However, in opposition to these results, El-Amrawy et al. found the Garmin Vivofit to be 97.01% accurate in calculating steps, but determined this during a 200, 500, and 1,000 step test in laboratory rather than free-living conditions [2].

Potential reasons for the conflicting step count results include the environment in which these values were taken. Our results match most closely to the experiment performed by Tully et al. most likely because both studies were conducted during a seven day period of free living with the Actigraph acting as the control device. The Garmin Vivofit most likely overestimates step count in these environments as intensity of the physical activity increases beyond the low level step test studied by El-Amrway et al. in the lab environment. In other words, there can be a difference between the devices with step count due to a free living environment differing from a controlled, lab environment. Usually in a free living environment, daily living activities require different intensities and movements that are not as precise in getting measurements by the devices as in a simple walk test. This can be seen in the fact that the controlled environment used in the study by El-Amrway et al. allowing for a more accurate step count measurement by the Garmin Vivofit in opposition to the free living environment results from Tully et al. as well as the present study. This suggests that the variability in the types of physical activity performed during a free living environment may be a negative factor in commercial activity monitors providing accurate step count measurements during daily living activities. Overall, this may mean that the technology in these devices are not as refined as the Actigraph when it comes to calculating step count.

A limitation of this study was the small sample size of participants. This was limiting in that it does not provide a fully representative population in that the participants were all college-aged students as well as skewed to having more female participants. Furthermore, another limitation of this study is that the free living environment is variable between each participant. This limits knowing how energy expenditure is measured based on different types of physical activities and different intensities of activity.

In conclusion, the Garmin Vivofit total energy expenditure results matched our expected results of having no significant difference between the Vivofit and the Actigraph. However, active energy expenditure and step results were unexpected with there being a significant difference between the Actigraph and Garmin Vivofit.

Manuscript References

1. Dannecker, Kathryn L., Sazonova, Nadezhda A., Melanson, Edward L., Sazonov, Edward S., and Browning, Raymond C. (2013). A comparison of energy expenditure estimation of several physical activity monitoring. *Medicine and Science in Sports and Exercise*, 2105-2112.
2. El-Amrawy, Fatema and Nounou, Mohamed I. (2015). Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthc Inform Res*, 21 (4), 315-320.
3. Ferguson, Ty., Rowland, Alex V., Olds, Tim, and Maher, Carol (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*, 12 (42), 1-9.
4. Lee, Jung-Min, Kin, Youngwon, and Welk, Gregory J. (2014). Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise*, 1840-1848.
5. Mifflin, M.D., St Jeor, S.T., Hill L.A. (1990). A new predictive equation for resting energy expenditure in healthy individuals. *American Journal of Clinical Nutrition, 51, 241-247.*
6. Nelson, Benjamin M., Kaminsky, Leonard A., Dickin, D. Clark, and Montoye, Alexander H. K. (2016). Validity of consumer-based physical activity monitors for specific activity types. *Medicine and Science in Sports and Exercise*, 1-36.
7. Tully, Mark A., McBride, Cairmeal, Heron, Leonnie, Allen, William, and Hunter, Ruth F. (2014). The validation of FitBit Zip physical activity monitor as a measure of free living physical activity. *BMC Research Notes*, 7 (952).

Table 1. Subject Characteristics

Characteristics	
Sex	8 females ; 2 males
Age	19.20 ± 1.23
Height (cm)	167.20 ± 11.44
Weight (kg)	62.430 ± 10.62
BMI	22.330 ± 3.08
% Total Body Fat	25.860 ± 8.30

Data are presented as means ± SD. BMI = body mass index

Figure 1. Mean Total Daily Energy Expenditure for 7 Days

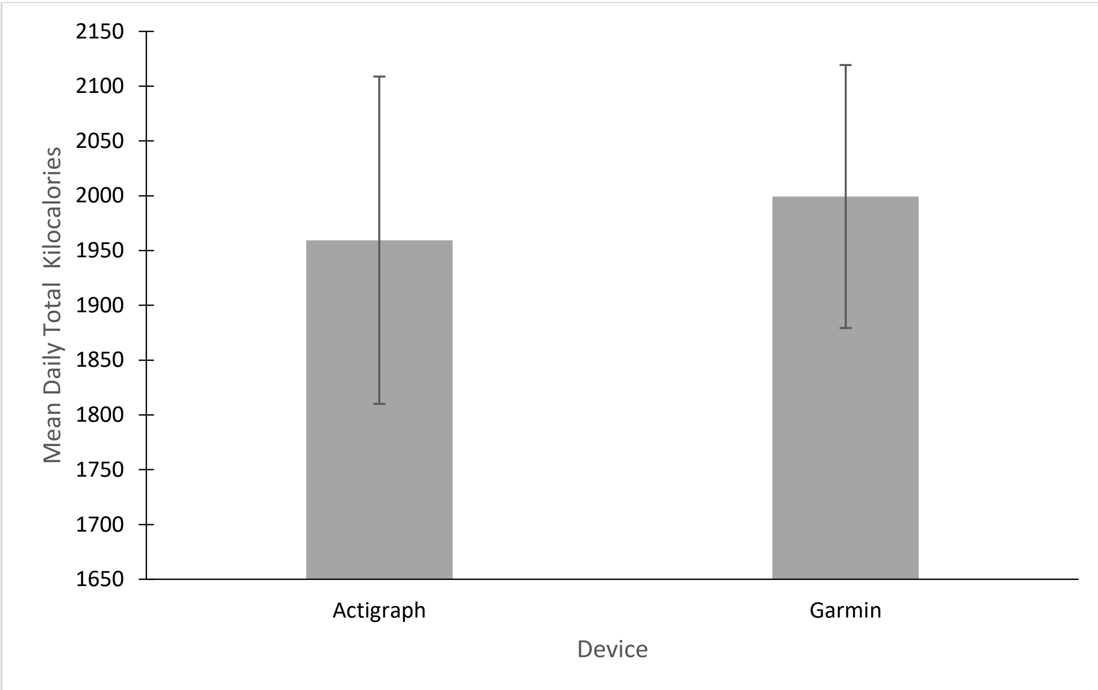
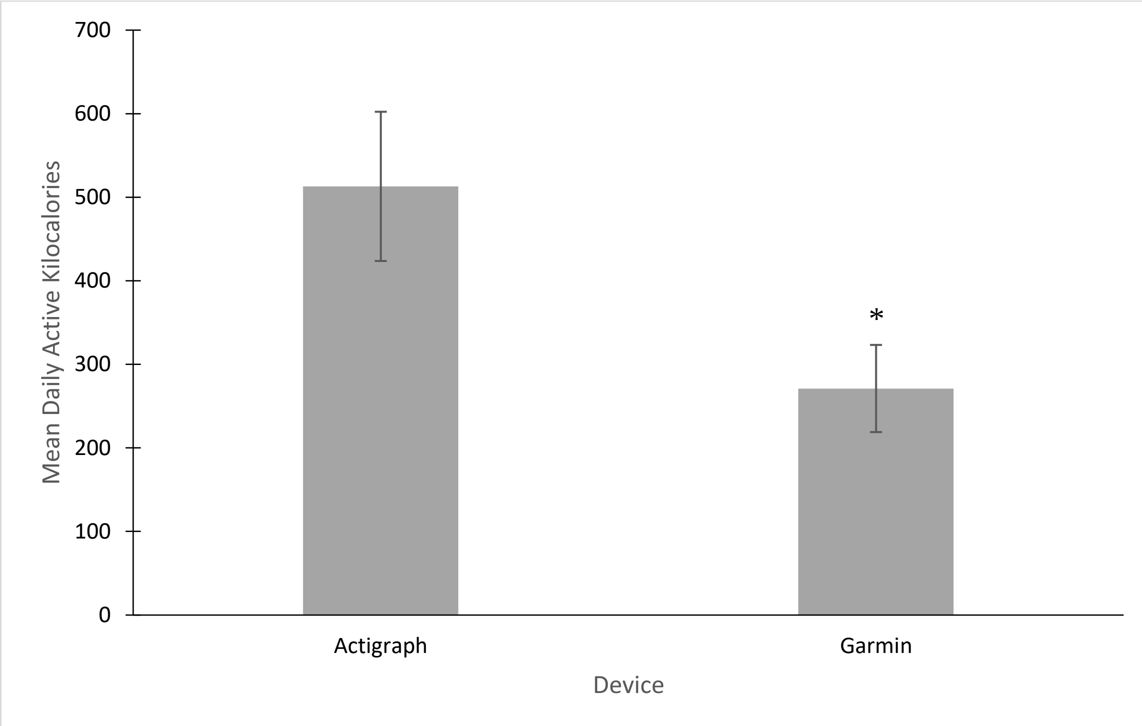
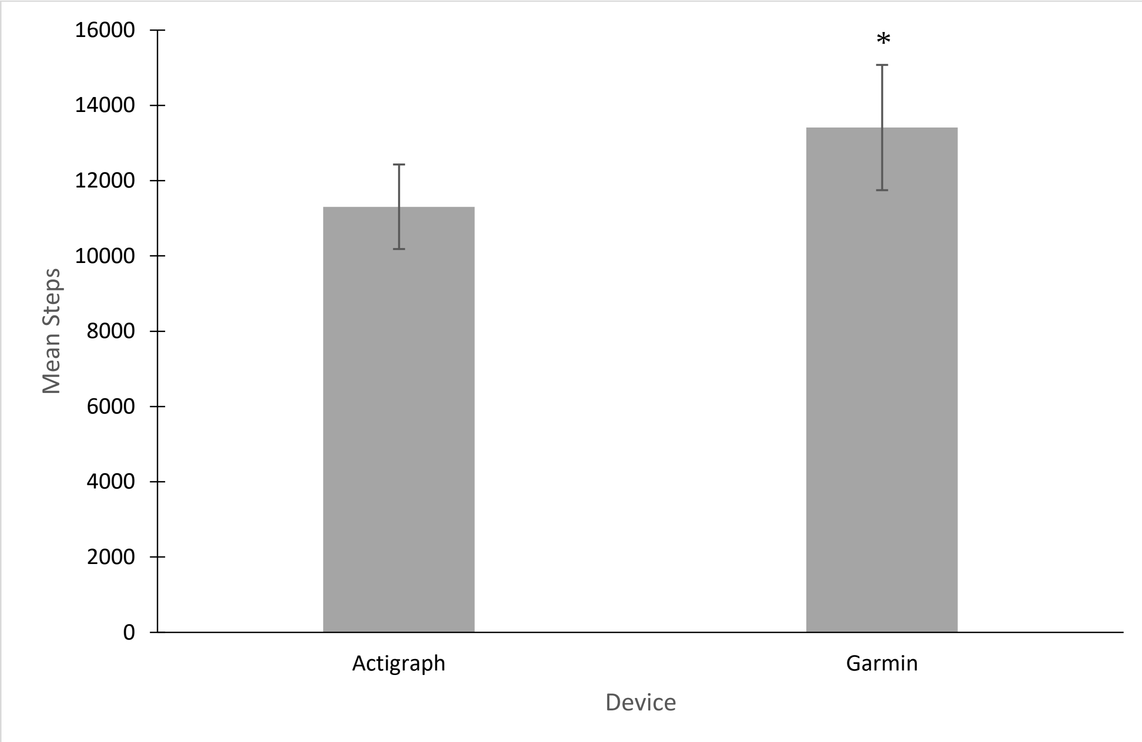


Figure 2. Mean Daily Active Energy Expenditure for 7 Days



*Indicates a significant difference from the Actigraph ($p < 0.05$)

Figure 3. Mean Daily Steps Averages for 7 Days



*Indicates a significant difference from the Actigraph ($p < 0.05$)

Figure 4. 7 Day Mean Total Kilocalories

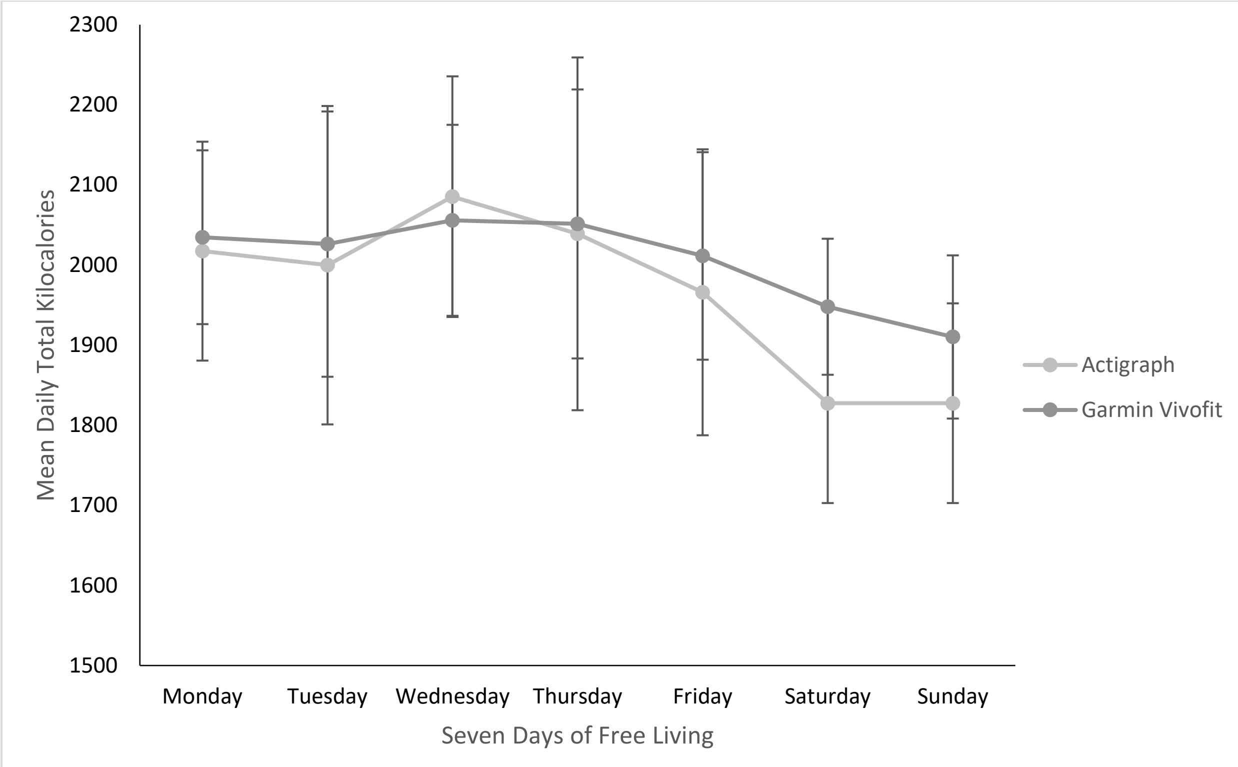
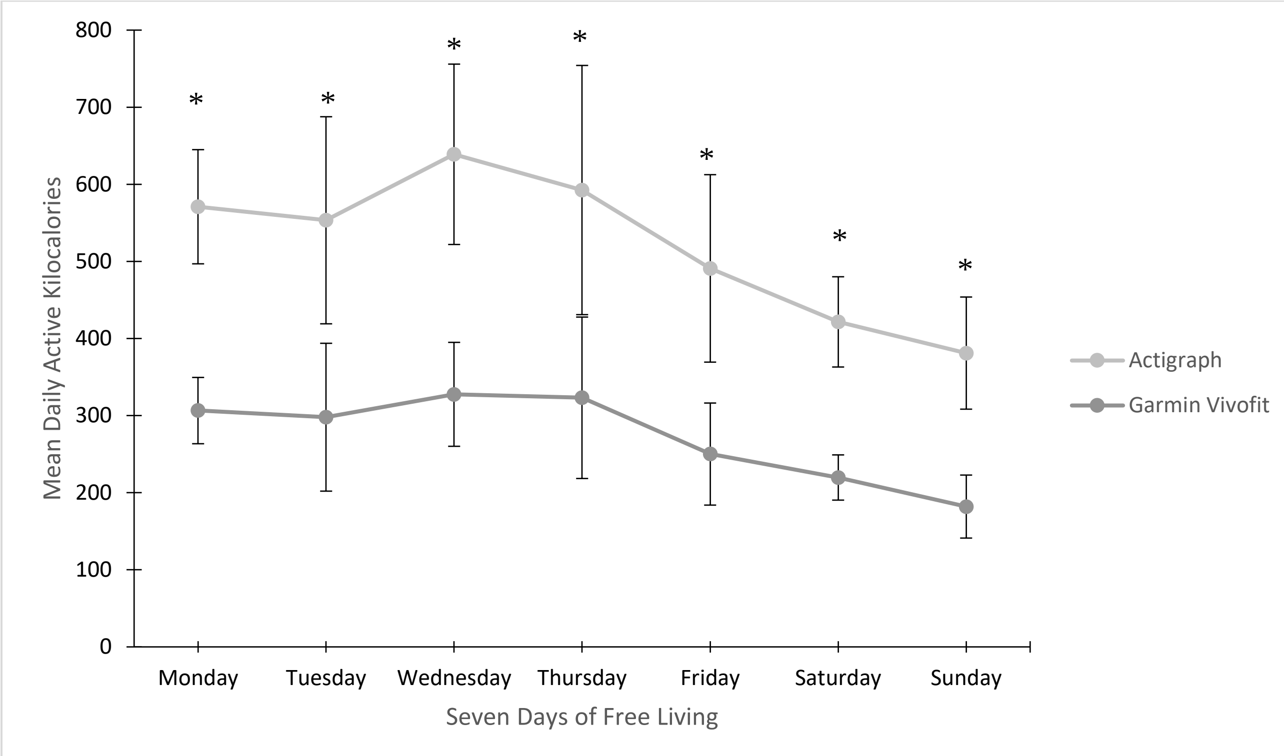


Figure 5. 7 Day Mean Active Kilocalories



*Indicates a significant difference from the Actigraph ($p < 0.05$)

Figure 6. Total Daily Energy Expenditure Correlation

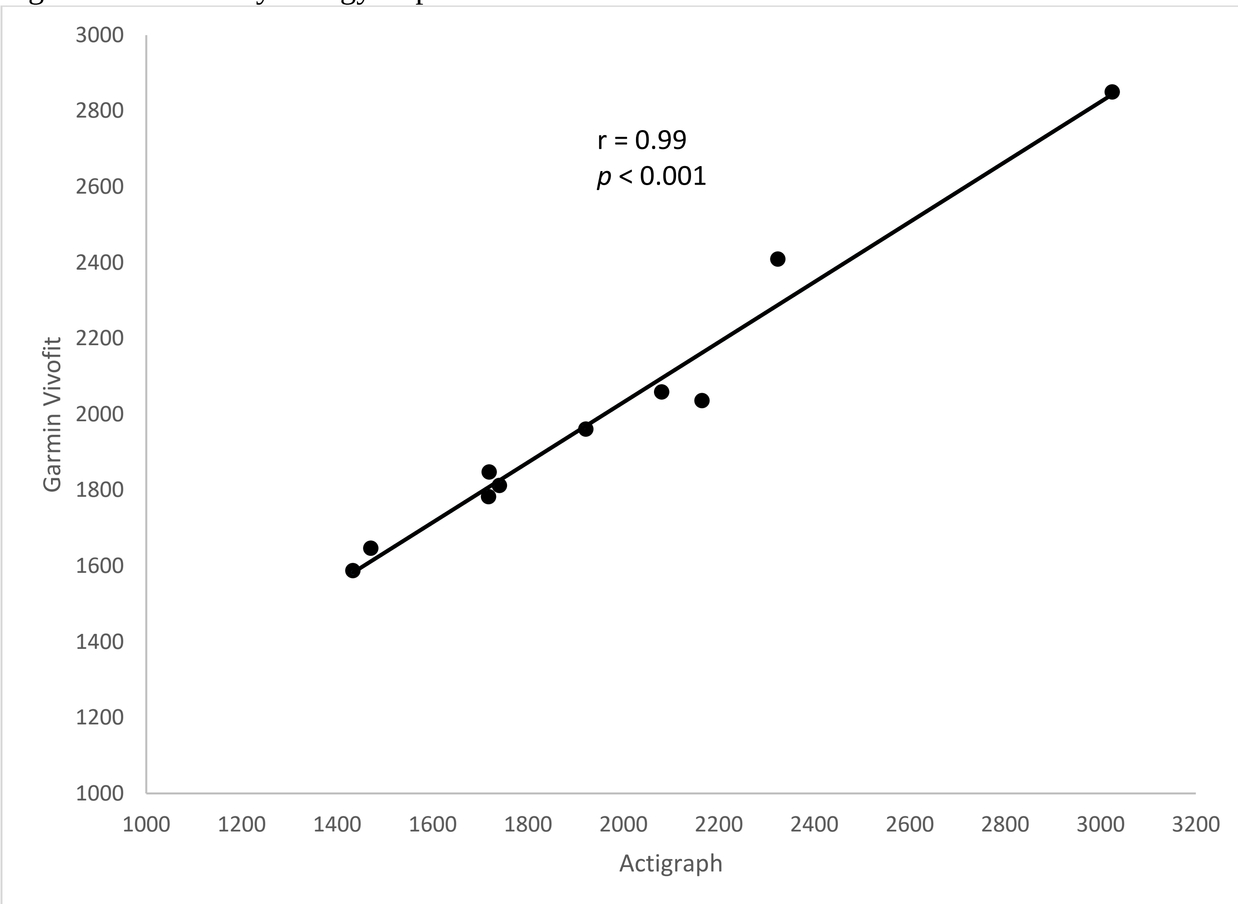
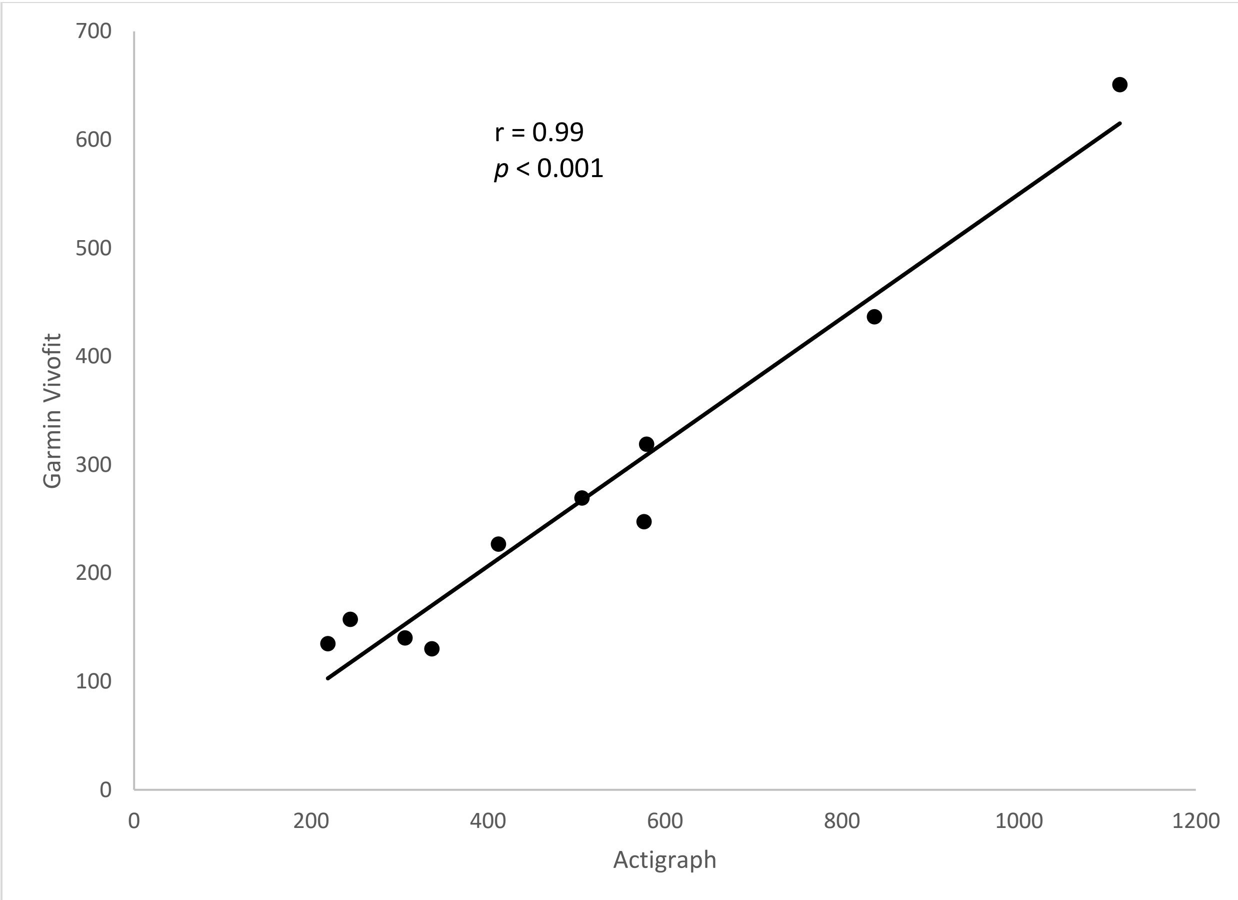


Figure 7. Daily Active Energy Expenditure Correlation



Appendix A

Email Recruitment Statement

Physical Activity Research at James Madison University

Do you own a Fitbit device or other similar wearable activity tracker? Do you wonder just how accurate they are? Researchers at James Madison University's Department of Kinesiology are conducting a study to assess the accuracy of commercially available physical activity and sleep monitors. We are looking or:

- Individuals 18 years and older
- Not currently smoking
- Are without significant heart, lung or metabolic disease
- Have no serious orthopedic or bone problems that prevent vigorous exercise

This study would require subjects to wear 3 activity trackers, 2 on the hip and 1 on the wrist, for 7 consecutive days while they are awake and during sleep. In addition, subjects will undergo body composition assessment, undergo a maximal treadmill exercise test, and a sub-maximal treadmill exercise trial over the course of 2 visits to the James Madison University Human Performance Laboratory. Subjects will receive a comprehensive report on their health status including body composition (% body fat), cardiorespiratory fitness, physical activity levels, and sleep quality. If you are interested in volunteering, please email the research staff at jmucpam@gmail.com

Appendix B

Informed Consent

James Madison University
Department of Kinesiology
Informed Consent

Purpose

You are being asked to volunteer for a research study conducted by Dr. Trent Hargens from James Madison University entitled, “The validity of commercially available physical activity monitors”

The goal of this study is to examine whether physical activity monitors that consumers can buy (the Fitbit Charge, the Fitbit Flex, or Garmin Vivofit) give the person accurate data when it comes to steps taken, energy expended (calories), and intensity of physical activity.

Experimental Procedures

You will be asked to visit the Human Performance Laboratory (HPL) in Godwin Hall 2 times over the course of about 8 – 10 days. Your total time commitment for participation in this study will be about 2 and a half hours (not counting the time you will be wearing devices while not at the HPL). You will be asked to wear 2 devices for a period of 7 days, one of which will be worn on your right hip, the other one will be worn on the wrist of the hand that you do not write with. These two devices will be worn during the day, as well as during the night when you sleep. The purpose of these two devices is to measure your physical activity level and sleep quality.

<u>Visit 1</u>

Upon completion of this informed consent, you will be asked to complete a short health history questionnaire providing information about your health and lifestyle characteristics. This will also help us to make sure you do not have any factors that may disqualify you from participation. You will then be asked to complete 2 additional questionnaires, one that asks about your ability to participate in physical activity, and one that asks about your current level of physical activity participation. Each of these questionnaires should take about 5 minutes to complete.

You will then have your height and weight measured, and then your body composition will be measured via a Dual-energy x-ray absorptiometer (DEXA). The DEXA scan will allow us to measure your percent body fat and percent lean mass. The DEXA is much like an x-ray machine. The DEXA will scan your entire body slowly, so you will need to lie on a table without moving for almost 10 minutes. You will feel no discomfort with this test.

Upon completion of the DEXA scan, you will be asked to complete a maximal treadmill exercise test. During the test, researchers will monitor your heart rate, blood pressure and your perceived exertion to the workload you will be doing. During the test, you will breathe only through a mouthpiece, with your nose clamped off, so that we can measure the amount of oxygen you use during exercise. The treadmill will start with a very slow speed and little to no grade, and will steadily increase in speed and grade throughout the test, so it will start easy, and become very intense. At the end of the test, it should be a best effort on your part. It may be as hard as any exercise you remember doing. The exercise test will last about 8 – 12 minutes. Prior to your arrival to the HPL that day, you will be asked to refrain from eating for 4 hours prior to your arrival, and to avoid caffeine and alcohol for that time period as well.

At the end of this visit you will be instructed on the proper procedures for wearing the 2 devices that you will then wear for 7 days in a row.

Visit 2

8 to 10 following Visit 1, you will be asked to return to the HPL. The day you return to the HPL you will be asked to not wear the two devices at all. For this visit you will be asked to complete a submaximal treadmill exercise trial while wearing 4 devices to monitor your physical activity and steps, the 3 your previously wore for 7 days, plus an additional 1 on your wrist. You will be asked to walk on the treadmill for a total of 20 minutes, with an additional 10 minutes for those who are able to tolerate treadmill running. The first 10 minutes will be fairly easy (5 minutes with no grade, 5 minutes with some grade), the second 10 minutes will be somewhat hard (5 minutes with no grade, 5 minutes with some grade), and the final 10 minutes (for those able) will be hard (5 minutes with no grade, 5 minutes with some grade). You will be monitored in a similar way as your maximal treadmill test (heart rate, blood pressure, oxygen measurement). You will be asked to again refrain from food, caffeine or alcohol for 4 hours prior to arrival at the HPL.

Risks

There are no risks associated with wearing a accelerometers (the physical activity monitoring devices). Also, there is no risk associated with heart rate, blood pressure, height, and weight. You will not be asked to change any of your personal habits during the course of the study. Measurements with associated risks include: the DEXA scan, the maximal treadmill exercise test, and the submaximal treadmill exercise trial.

The amount of radiation that you will receive in the DEXA scan is less than the amount you will receive during a transatlantic flight, and is equal to about 1/20 of a chest x-ray. If you think that you may be pregnant, please inform the research staff immediately, as the radiation from the DEXA could potentially harm the fetus. If you choose not to complete the DEXA scan, you will not be able to continue with the study.

There is a risk of abnormal changes during the maximal and submaximal treadmill exercise tests. These changes may include abnormal blood pressure, fainting, heart rhythm disorders, stroke, heart attack, and death. The chance of serious heart problems during maximal exercise among adults is very small (less than 1/10,000 maximal exercise tests). Every effort will be made to minimize risks of an abnormal response by reviewing you health history and providing adequate supervision of the exercise test. All staff are certified by the American Heart Association in BLS (Basic Life Support), and all tests will be supervised by an individual certified by the American College of Sports Medicine.

Benefits

Participation may include knowledge about your health status. You will receive information on your percent body fat, an assessment of your sleep quality, an assessment of your physical activity level, and cardiovascular fitness. Indirect benefits of participating in this study will be helping the researchers better understand if commercially available physical activity and sleep monitors are accurate in the data they provide the consumer.

Inquiries

If you have any questions or concerns or you would like to receive a copy of the final aggregate results of this study, please contact Dr. Trent Hargens at hargenta@jmu.edu or (540) 568-5844.

Questions about Your Rights as a Research Subject
Dr. David Cockley
Chair, Institutional Review Board
James Madison University
(540) 568-2834
cocklede@jmu.edu

Confidentiality

All data and results will be kept confidential. You will be assigned an identification number. At no time will your name be identified with your individual data. The researcher retains the right to use and publish non-identifiable data. All paper data will be kept secured in a locked cabinet in a locked office. All electronic data will be kept on a password-protected computer in encrypted file folders. Final aggregate results will be made available to participants upon request.

Freedom of Consent

Your participation is entirely voluntary. You are free to choose not to participate. Should you choose to participate, you can withdraw at any time without consequences of any kind.

I have read this consent form and I understand what is being requested of me as a participant in this study. I freely consent to participate. I have been given satisfactory answers to my questions. The investigator provided me with a copy of this form. I certify that I am at least 18 years of age. By clicking "Yes" to the question below and submitting this confidential online survey, I am consenting to participate in this research.

Do you provide consent to participate in the research study entitled, "The validity of commercially available physical activity monitors"?

- ○ Yes
- ○ No

Please enter your name here

Please enter today's date

Bibliography

1. Craig, Cora L., Marshall, Alison L., Sjostrom, Michael, Bauman, Adrian E., Booth, Michael L., Ainsworth, Barbara E., Pratt, Michael, Ekelund, Ulf, Yngve, Agneta, Sallis, James F., and Oja, Pekka (2003). International physical activity questionnaire: 12 country reliability and validity. *Medicine and Science in Sports and Exercise*, 1381-1390.

2. Dannecker, Kathryn L., Sazonova, Nadezhda A., Melanson, Edward L., Sazonov, Edward S., and Browning, Raymond C. (2013). A comparison of energy expenditure estimation of several physical activity monitoring. *Medicine and Science in Sports and Exercise*, 2105-2112.

3. Diaz, Keith M., Krupka, David J., Chang, Melinda J., Peacock, James, Ma, Yao, Goldsmith, Jeff, Schwawrtz, Joseph E., and Davidson, Karina W. (2015). Fitbit: an accurate and reliable devices for wireless physical activity tracking. *International Journal of Cardiology*, 185, 138-140.

4. El-Amrawy, Fatema and Nounou, Mohamed I. (2015). Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthc Inform Res*, 21 (4), 315-320.

5. Ferguson, Ty., Rowland, Alex V., Olds, Tim, and Maher, Carol (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*, 12 (42), 1-9.

6. Kooiman, Thea J.M., Dontje, Manon L., Sprenger, Siska R., Krijnen, Wim P., P. van der Schans, Cees, and Groot, Martijn de Groot (2015). Reliability and validity of ten

consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation*, 7 (24), 1-12.

7. Lee, Jung-Min, Kin, Youngwon, and Welk, Gregory J. (2014). Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise,* 1840-1848.

8. Mifflin, M.D., St Jeor, S.T., Hill L.A. (1990). A new predictive equation for resting energy expenditure in healthy individuals. *American Journal of Clinical Nutrition, 51, 241-247.*

9. Nelson, Benjamin M., Kaminsky, Leonard A., Dickin, D. Clark, and Montoye, Alexander H. K. (2016). Validity of consumer-based physical activity monitors for specific activity types. *Medicine and Science in Sports and Exercise*, 1-36.

10. Noah, J. Adam, Spierer, David K., Gu, Jialu, and Bronner, Shaw (2013). Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. *Journal of Medical Engineering and Technology,* 37 (7), 456-462.

11. Pescatello, Linda S. (2014). *ACSM's guidelines for exercise testing and prescription* (9th ed.). Baltimore, MD: Lippincott Williams & Wilkins.

12. Stackpool, Caitlin M., Porcari, John P., Mikat, Richard P., Gilette, Cordial, and Foster, Carl (2014). The accuracy of various activity trackers in estimating steps taken and energy expenditure. *Journal of Fitness Research*, 3 (3), 32-48.

13. Storm, Fabio A., Heller, Ben W., and Mazza, Claudia (2015). Step detection and activity recognition accuracy of seven physical activity monitors. *PLoS ONE*, 10 (3), 1-13.

14. Takacs, Judit, Pollock, Courtney L., Guenther, Jerrad R., Bahar, Mohammadreza, Napier, Christopher, and Hunt, Michael A. (2014). Validation of the FitBit One activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport*, 17, 496-500.

15. Tully, Mark A., McBride, Cairmeal, Heron, Leonnie, Allen, William, and Hunter, Ruth F. (2014). The validation of FitBit Zip physical activity monitor as a measure of free living physical activity. *BMC Research Notes,* 7 (952).